Research Paper | Statistics | E-ISSN No : 2454-9916 | Volume : 3 | Issue : 7 | July 2017

NONPARAMETRIC POST HOC TEST WITH ADJUSTED P VALUE

Mital C. Shingala [1] | Dr. Arti Rajyaguru [2]

[1] Assistant Professor, Shree J.D.G. Commerce & Shree S.A.S. College of Management, Varachha, Surat-395006, Gujarat, India.
[2] Professor, Department of Statistics, Veer Narmad South Gujarat University, Udhana Magdalla Road, Surat- 395007, Gujarat, India.

ABSTRACT

In this paper we have compared Non parametric post hoc tests with adjusted p value for easy application purpose.

KEYWORDS: Multiple Comparisons, Non Parametric Post Hoc Test, Adjusted P value

1. INTRODUCTION:

When data consist of one nominal variable and one measurement variable, usually one way ANOVA is used but when the measurement variable does not meet the normality assumption of a one-way ANOVA then parametric method is not applicable and when original data set actually consists of one nominal variable and one ranked variable; we cannot apply ANOVA. The non-parametric techniques which have been developed for k sample problem require no assumptions beyond continuous populations and therefore it is applicable under any circumstances.

One of the assumptions of the parametric analysis is that the variability is approximately the same across all groups. If this assumption does not hold then researcher should first try to transform the response variable, perhaps using a log or square root transformation. Hopefully this will be stabilizing the variance across the groups. However, in certain situations none of the transformation resolves this problem. In this situation also researcher should consider using a non parametric test (Newman, 1995).

If the normality assumption is violate or the sample sizes from each of the k populations are too small to assess normality, Kruskal Wallis (KW) test is used to compare the distribution of different populations. Kruskal Wallis (KW) test is the non parametric equivalent to the omnibus F test in a one way ANOVA (which is used with matrix dependent variable). KW test is used when the dependent variable consist of ranks. It tests the null hypothesis that the location of each group is the same in the population. If the null hypothesis is rejected, then at least one of the locations is different from the others. When the KW test is significant, perform follow up pair wise tests.

It is important to realize that Kruskal-Wallis test is an omnibus test that enables to test the general hypothesis that all population medians are equal but cannot tell which specific groups of independent variable are statistically significantly different from each other; it only tells that at least two groups are different. But the researcher is not just interested in this general hypothesis but in comparisons amongst the individual groups. Since we may have three, four, five or more groups in our study design, determining which of these groups differ from each other is important and then post hoc test is used

There are two ways to apply non-parametric post hoc procedures, the first being to use Mann Whitney tests. However, if we use lots of Mann Whitney tests, Type I error rate will inflate, therefore not preferable. However, if we want to use lots of Mann–Whitney tests to follow up a Kruskal–Wallis test, we can if we make some kind of adjustment to ensure that the type I errors don't build up to more than .05. The easiest method is to use a Bonferroni correction, which in its simplest form just means that instead of using .05 as the critical value for significance for each test, you use a critical value of .05 divided by the number of tests conducted.

So in this research paper, we will discussed and compare Bonferroni method and modification of it like Holm, Hochberg, Hommel, Holland and Rom. All this methods are based on adjusted p value.

An adjusted p value is defined as the smallest significance level for which the given hypothesis would be rejected, when the entire family of tests is considered. The decision rule is to reject the null hypothesis when the adjusted p-value is less then α; in most cases, this procedure controls the FWE at or below α level.

In this research paper, we have discussed tests based on adjusted p values such that, if the adjusted p value for an individual hypothesis is less than the chosen significance level α, then the hypothesis is rejected with FWE not more than α. It includes Bonferroni procedure and modification of that procedure by Holm, Holland & Copenhaver, Hommel, Hochberg and Rom. From them some of the methods are Single step procedure and others are sequential methods. Further Sequential methods can be categorized in two ways i.e. Step up method and step down method.

Single Step/Simultaneous Procedure:

It is also called Single Step (SS) procedure. The single step procedure sets a single criterion for testing all individual hypotheses. SS procedure conducts all comparisons regardless of any other comparison is significant or not using a constant critical value (Einot & Gabriel, 1975). SS procedures are valid to use both for hypothesis testing and to calculate confidence intervals.

In SS procedure, the decision about any hypothesis Hi does not depend on the decision about any other hypothesis Hj therefore the hypotheses can be tested without reference to one another.

Sequential Procedure:

It is also called Step Wise (SW) procedure. A step wise procedures consider either the significance of the omnibus test or the significance of other comparisons or both in evaluating the significance of a particular comparison.

In SW procedure, the hypotheses are tested in a specific order, generally determined by the magnitudes of the test statistics or the associated p-values, pi and the decisions on them are made in a stepwise manner. The decisions on the earlier hypotheses in the order may affect those on the later hypotheses in the order.

A major disadvantage of stepwise (multi-stage) method is that it does not allow the construction of confidence interval, which is extremely useful for the interpretation of the results.

SW procedures can be further subdivided into 2 categories.

i. Step Up procedures (SD)

ii. Step Down procedures (SU)

Step Up procedures:

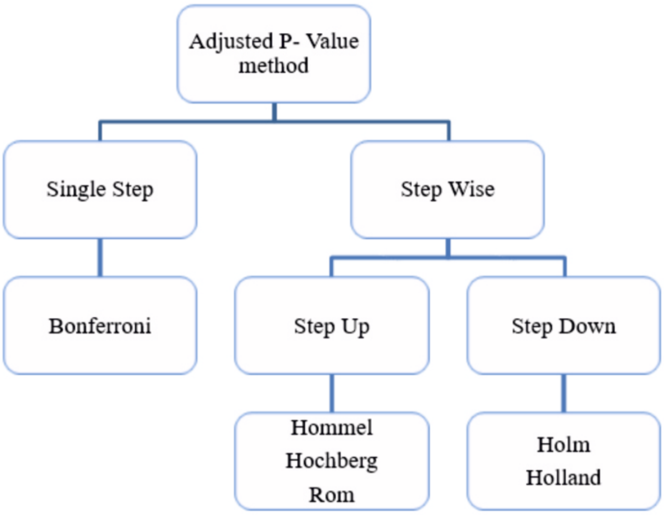
In SU procedure, the hypotheses are tested beginning with the least significant one and testing continues until a hypothesis is rejected at which point all the remaining hypotheses are rejected by implication without actually testing them (Tamhane & Dunnett, 1999).

SU procedure begins by testing all minimal hypotheses and then steps up through the hierarchy of implying hypotheses. If any hypothesis is rejected, then all of its implying hypotheses are rejected without further tests; thus a hypothesis is tested if and only if all of its implied hypotheses are retained.

Step Down Procedure:

A step down procedure begins by testing the overall intersection hypothesis and then steps down through the hierarchy of implied hypotheses. If any hypothesis is not rejected, then all of its implied hypotheses are retained without further tests; thus a hypothesis is tested if and only if all of its implying hypotheses are rejected. In SD Procedure, the hypotheses are tested beginning with the most significant one and testing continues until a hypothesis is not rejected at which point all the remaining hypotheses are accepted by implication without actually testing them.

2. Non parametric Tests:
These methods can be classified on the basis of the either simultaneous or sequential procedure.



2.1. Bonferroni Test (1961):
The Bonferroni method applies to both continuous and discrete data. This method is flexible because it controls the FWE for tests of joint hypotheses about any subset of m separate hypotheses (including individual contrasts). The procedure will reject a joint hypothesis H0 if any p-value for the individual hypotheses included in H0 is less than α/c. Bonferroni method, however, yields conservative bounds on Type I error hence it has low power. This procedure controls the FWE at α without any further assumption on the dependence structure of the p value. This method is design for comparisons involving pair wise comparisons as well as combinations of means, provided the number of comparisons to be made is fixed in advance. It is recommended for non-orthogonal contrast because it splits the type I error rate equally among all comparisons.

The Bonferroni procedure is used for evaluating a small number of contrast that is selected prior to observing the data while preserving a selected family wise type I error rate (Ott & Longnecker, 2010). The researcher must have sufficient theory about the phenomena of interest in order to know which contrasts to specify. It is appropriate when the number of comparisons exceeds the number of degrees of freedom between groups. This method controls type I error but it will increase type II error. The purpose of Bonferroni procedure is to reduce the probability of identifying significant results that do not exist, that is, to guard against making Type I error in the testing process (Mchugh, 2011). These potential for error increases with an increase in the number of tests being performed in a given study and is due to the multiplication of probabilities across the multiple tests (Mchugh, 2011).

Test statistics:

$$t_{cal} = \frac{\left|\bar{X}_i - \bar{X}_j\right|}{\sqrt{MS_w\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}} \qquad \text{...(1)}$$

Where,
Ms_w = Error mean square from the ANOVA.

$\bar{X}_i$ and $\bar{X}_j$ are the two means being compared.

n_i and n_j are the respective sample sizes from population i and j.

Critical Value:

$$t_{tab} = t_{\alpha/2c}, df_w \qquad \text{...(2)}$$

Where,
t is the value from the t distribution.

c=number of pair wise comparisons in the family.

For complex comparison:

$$t_{cal} = \frac{\left|\sum c_i \bar{X}_i\right|}{SE} \qquad SE = \sqrt{MS_w\left(\sum \frac{c_i^2}{n_i}\right)} \qquad \text{...(3)}$$

Decision procedure:
Reject the null hypothesis if $t_{cal} \geq t_{tab}$; do not reject H_0 otherwise.

Confidence Interval:

$$\bar{X}_i - \bar{X}_j \pm t'_{\alpha/2,c,df_w}\sqrt{MS_w\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \qquad \text{...(4)}$$

Here, margin of error depends on the number of comparisons.

Confidence interval for the contrasts:

$$\sum_{i=1}^{k} c_i\bar{x}_i \pm t'_{\alpha/2,c,df_w}\sqrt{MS_w\sum_{i=1}^{k}\frac{c_i^2}{n_i}} \qquad \text{...(5)}$$

Advantages:
1. This method is highly flexible because it can be applied to test any subset of hypotheses for continuous, discrete data and even correlated tests.
2. This method is used for any design.
3. When number of comparisons are small (i.e. number of comparison less than or equal to number of groups-1) it gives smaller confidence interval.
4. This method is useful in confirmatory research when a family of selected pair wise comparisons is specified prior to data collection; it reduces the problem of alpha inflation (Mchugh, 2011).
5. Bonferroni method can also test complex pairs. (Mchugh, 2011).
6. This method has relatively good power for small sets of planned comparison (Toothaker, 1993).

Disadvantages:
1. It cannot be used for data snooping because the tests of interests are specified prior to data analysis.
2. This method has lower power to reject an individual hypothesis and it lacks power if several highly correlated tests are undertaken (Simes, 1986; Li, 2009; Hommel, 1988).
3. It has power to quickly decline as the number of comparisons increases (Olejnik et al., 1997; Toothaker, 1993).
4. It is not a tool for exploratory data analysis.
5. The test does not take into account whether the findings are consistent with theory and past research. If consistent with previous findings and theory, an individual result should be less likely to be a Type I error.

2.2. Holm Test (1979):
It is a modification of Bonferroni procedure that yields a more powerful test. The goal of Holm method is to increase the power of the statistical tests while keeping under control the FWE. It is a step down procedure. It is also called a sequential rejection method because it examines each hypothesis in an ordered sequence and the decision to accept or reject the null hypothesis depends on the results of the previous hypothesis tests (Tamhane et al, 1998). He was the first to formally introduce a sequentially rejective Bonferroni procedure. Bonferroni method does not account for the correlations between the test statistics, the Holm procedure can be improved.

Holm method can be applied to almost any data because of its non-parametric nature. This test can be applied in any pair wise comparison where the classical Bonferroni test is usually applied. It is applicable when pair wise comparisons of median or linear combinations or non linear combinations of median are used. It is used to perform priori comparison. For several a priori contrasts, not necessarily pair wise, it controls FWE while at the same time maximizes the power (Howell, 2007).

Assumptions:
There are no restrictions on the type of test; the only requirement is that it should be possible to calculate the obtained level for each separate test. Further, there are no problems to include in the analysis only for the a priori interesting hypotheses, while more special multiple tests usually include on all hypotheses of a certain kind.

Holm's procedure may be used either as a protected test or as an unprotected test but the protected version is preferred due to the additional power gains. But when there exist logical implications among the hypotheses, problems arise which we have to take in to consideration (Holm, 1979). So, Holm's procedure makes no distributional assumptions, logical assumptions about the hierarchy of the hypotheses to be tested and does not assume independence of comparisons

(Zweifel, 2014).

Procedure:

Order the p values, $p_{(1)} \geq \ldots\ldots \geq p_{(c)}$, and denote the corresponding hypotheses, $H_{(1)},\ldots..,H_{(c)}$. Start with the smallest p value, $p_{(c)}$. If $p_{(c)} > \alpha/c$, then stop testing and accept all the hypotheses; otherwise reject $H_{(c)}$ and go to the next step. In general, if testing has continued to the i_{th} step ($1 \leq i \leq c$) and if $p_{(c-i+1)} > \alpha/(c-i+1)$, then stop testing and accept all the remaining hypotheses, $H_{(c-i+1)},\ldots.., H_{(1)}$; otherwise reject $H_{(c-i+1)}$ and go to the next step.

In short, this procedure rejects the specific hypothesis $H_{(I)}$ for i = 1,2,…,c, provided both $P_{(i)} \leq \alpha/(c-i+1)$ and $H_{(1)},\ldots, H_{(i-1)}$ have all been rejected.

Like Bonferroni procedure, Holm's procedure can also modify p-values directly multiplying the p-value by the adjusted C-*i*+1, where *i* is an index of the step associated with the p value.

For unequal sample size, the test statistics is same as Bonferroni given by….(1)

For equal sample size, the test statistics is given as

$$t' = \frac{\left|\bar{x}_i - \bar{x}_j\right|}{\sqrt{\dfrac{2MS_{error}}{n}}} \quad \ldots(6)$$

Calculate t' for all contrasts of interest and then arrange the *t'* values in increasing order without regard to sign. This ordering can be represented as $|t'_1| \leq |t'_2| \leq |t'_3| \leq \cdots \leq |t'_c|$, where c is the total number of contrasts to be tasted.

The first significance test is carried out by evaluating t_c against the critical value in Dunn's table corresponding to *c* contrasts. In other words, t_c is evaluated at α' = α/c. If this largest *t'* is significant, then we test the next largest *t'* (i.e. $t'_{(c-1)}$) against the critical value in Dunn's table corresponding to *c*-1 contrasts. Thus, $t'_{(c-1)}$ is evaluated at α'=α/(*c*-1). The same procedure continues for $t'_{(c-2)}$, $t'_{(c-3)}$, $t'_{(c-4)}$,…until the test returns a non-significant result. At that point we stop testing. Holm has shown that such a procedure continues to keep FWE ≤ α, while offering a more powerful test.

The logic behind the test is that when we reject the null for t_c, we declare that null hypothesis to be false. If it is false, that only leaves *c*-1 possibly true null hypotheses, and so we only need to protect against *c*-1 contrasts. A similar logic applies as we carry out additional tests. This logic makes particular sense when even before the experiment is conducted we know that some of the null hypotheses are almost certain to be false. If they are false, there is no point in protecting from erroneously rejecting them.

Critical Value:

$$\alpha/_{c-i+1} \quad \ldots(7)$$

Decision procedure:

Reject $H_{(1)}$ to $H_{(i-1)}$ if

$$P_{(i)} \leq \frac{\alpha}{c-i+1} \quad \ldots(8)$$

α will change at all stages because of its step down nature.

The critical value of this method is based on the Bonferroni inequality.

Advantages:

1. This method is flexible and simple to implement.

2. It controls the FWE in the strong sense, i.e. it guarantees control of generalized Type I error probability to be at most α (Hochberg, 1988; Schochet, 2008; Ekenstierna, 2004; Hochberg & Benjamini, 1990; De Muth, 2006).

3. This archives lower type II error while keeping the type I error rate at level less than α (Hochberg & Benjamini, 1990).

4. It can be used for equal as well as for unequal sample size.

Disadvantages:

1. Power of this method is small If all the hypotheses are almost true but it may be considerable if a number of hypotheses are completely wrong (Holm, 1979).

2. It does not compute confidence intervals.

3. It does not consider the logical interrelationships among the c hypothesis.

4. It becomes very conservative when the numbers of comparisons are large and when tests are not independent (De Muth, 2006).

2.3. Holland & Copenhaver Test (1987):

It uses the Sidak (1967) inequality to set the criterion for each hypothesis test. It is a step down procedure. When there is need for further research in situations, where there is no logical inter relationship among the hypotheses, this method is useful.

Assumptions:

Positive orthant dependence of the test statistics is satisfied.

Procedure:

Let $p_{(1)},\ldots,p_{(c)}$ be the ordered p values (smallest to largest) and $H_{(1)},\ldots,H_{(c)}$ be the corresponding hypotheses. Suppose i is the smallest integer from 1 to m such that $p(i) > 1 - (1 - \alpha)^{1/(c-i+1)}$; the Holland-Copenhaver procedure rejects $H_{(1)}$ to $H_{(i-1)}$ and retains $H_{(i)}$ to $H_{(c)}$ (Olejnik et al, 1997).

Test Statistics:

For unequal sample size, the test statistics is same as Bonferroni given by …(1)

For equal sample size, the test statistics is same as Holm given by …(6)

Critical Value:

$$1 - (1 - \alpha)^{1/(c-i+1)} \quad \ldots(9)$$

Decision procedure:

Reject $H_{(1)}$ to $H_{(i-1)}$ if

$$p(I) < 1 - (1 - \alpha)^{1/(c-i+1)} \quad \ldots(10)$$

Advantages:

This method is conservative under the condition that the test statistics are positive orthant dependent.

Disadvantages:

Applicability of this method is slightly less than the Holm procedure because of the requirement of positive orthant dependent condition for test statistics.

2.4. Hommel Test (1988):

Hommel (1988) employs the closure principle to extend Simes test and developed a stepwise multiple testing procedure controlling FWE. It is based on the Simes (l986) equality. This is a step up method and it is protected test. This procedure is conservative only when the test statistics are independent, because it based on the Simes equality for independent p values. It is not always necessary to test every possible combination of hypothesis i.e. it can also be used for few comparisons.

The work of Hommel's who generalized Simes procedure that it gives strong control of FWE whenever Simes original procedure does achieve weak control (e.g. with independent tests).

Assumptions:

Test statistics are independent.

Procedure:

Reject all hypothesis that have a p value ≤ α/j' where j is defined as

$$j = \max\left(i' \in \{1,\ldots c\} : p_{(c-i'+k)} > \frac{k\alpha}{i'} \; for\; k = 1,\ldots i' \right) \quad \ldots(11)$$

If j is non empty, reject H_i whenever $P_i \leq \alpha/j'$ with j'=max j. If j is empty, reject all H_i (i=1,2,…c)

This procedure includes two stages. The first stage uses the obtained p-values to compute the number of members in J. The second stage obtains the significance level of rejection using α'=α/j', where j' is the largest number in J.

Test statistics is same as Holm given in…(6)

Critical Value

$$\alpha/j' \quad \ldots(12)$$

Advantages:

1. It protects the FWE only when test statistics are independent (Dmitrienko et al, 2009; Olejnik et al, 1997).

2. The uniqueness of the Hommel procedure is that it not only considers the order of the tests but also takes the obtained p values into the calculation while computing the α'.

Disadvantages:

1. This method is relatively complicated.

2. When correlations between variables are negative, the test can sometimes

allow slightly more Type I errors than the stated maximum family wise error.

2.5. Hochberg Test :(1988)

It is a modification of Dunn procedure. This procedure uses critical values identical to those used in Holm procedure but provides a potential for increased power by conducting the tests in a step-up rather than step down sequence. It is a step up method and based on the Simes (1986) equality.

Assumptions:

Tests are independent of one another.

Procedure:

Hochberg derived an even sharper procedure which uses the ordered p_is but in a different way from Holm's procedure. This procedure starts by examining the largest p-value p(c). If p(c) ≤ α, then H(c) and all other hypotheses are rejected. If not, H(c) is not rejected and one proceeds to compare $p_{(c-1)}$ with α/2. If the former is smaller, then $H_{(c-1)}$ and all hypotheses with smaller p-values are rejected. Generally, one proceeds from highest to lower p-values, retaining Ho, if its p-value satisfies p(i) > α/(c — i + 1). One stops the procedure at the first ordered hypothesis when that inequality is reversed. This hypothesis is rejected and all hypotheses with lower or equal p-values. This is always a sharper procedure than Holm's.

Critical Value:

$$\alpha / c - i + 1 \qquad \ldots(13)$$

Decision procedure:

Reject $H_{(1)}$ to $H_{(i)}$ for any i=c,c-1,….1 if

$$P_{(i)} \leq \frac{\alpha}{c-i+1} \qquad \ldots(14)$$

Advantages:

1. This procedure has strong control over the FWE α even if the free combination condition is not satisfied (Holm, 1979; Holland & Copenhaver, 1987; Olejnik et al, 1997).
2. It controls the FWE under the same conditions for which the Simes global test control the Type I error rate.
3. This method always achieves the same type I FWE control and lower type II error rates (Hochberg & Benjamini, 1990).
4. It has nice characteristic that no adjusted p value can be larger than the largest of the unadjusted P values (Wright, 1992).
5. This method is able to reject at least one individual hypothesis when the global null hypothesis is rejected. This property of consonance makes Hochberg procedure easy to interpret (Rom, 1990).

Disadvantages:

1. It lacks the stability under certain conditions, for example, when the test statistics are dependent or correlated (Schochet, 2008).
2. It only can be applied in the independent hypotheses tests (Olejnik, et al. 1997; Schochet, 2008).

2.6. Rom Test (1990):

It is a modification of Hochberg procedure to increase the statistical power. It is a step up procedure. Increased power is achieved by identifying the appropriate adjusted significance levels that control the Type I error rate at exactly the nominal level when test statistics are independent (Olejnik et al, 1997).

Assumptions:

Test statistics are independent.

Procedure:

The Rom procedure differs from the Hochberg procedure when the adjusted significance level is obtained. Both procedures set $\alpha'_{(m)}$ equal to α and $\alpha'_{(m-1)}$ equal to α/2, but the remaining m - 2 adjusted significance levels differ. The adjusted significance levels are determined recursively as

$$\alpha'_{m-i+1} = \left[\sum_{j=1}^{i-1} \alpha^{j} - \sum_{j=1}^{i-2} \binom{i}{j} \alpha'^{\,i-j}_{(m-j)} \right] \Big/ i \qquad \ldots(15)$$

i=1,2,…m

where $\alpha_{k-1} = \alpha$ and $\alpha_{k-2} = \alpha/2$.

It is step up procedure with different critical value of $c_1=\alpha$, $c_2=\alpha/2$, $c_3=\alpha/3 + \alpha^2/12$ etc.

First, we denote $H_{(1)}$ as the hypothesis with the largest p-value and $H_{(m)}$ as the hypothesis with the smallest p-value.

The testing starts by comparing $p_{(1)}$ with $\alpha_{(1)}$ and stops when $p_{(i)} < \alpha_{(i)}$. Then $H_{(1)}$ to $H_{(i+1)}$ retained and $H_{(i)}$ to $H_{(m)}$ rejected. The computing equation for solving α_i's can be divided into three parts.

The first part is $\alpha^1+\alpha^2+\ldots\alpha^{i-1}$ and the second part is $\binom{i}{1}(\alpha_{(2)}^{\,i-1})+\binom{i}{2}(\alpha_{(3)}^{\,i-2})+\ldots+\binom{i}{i-2}(\alpha_{(i-1)}^{\,2})$

The third part is to solve for α_i', which subtract the second part from the first part, and divide the difference by I.

Advantages:

1. It exactly controls the FWE at α for independent test statistics (Schochet, 2008).
2. It gives motivation of lowering type II error.
3. The Rom procedure having the desired FWE only for independent test, for complex comparison.

Disadvantages:

1. The calculation of this method is complicated and iterative.
2. It provides adjusted critical values for up to 10 tests when the overall alpha equals 0.05 and 0.01. The numbers of hypothesis test increases, the calculations become impractical even when a computer is used.

3. Comparison:

The methods discussed above are compared with respect to different aspects like Conservatism, Power and Confidence Interval estimation and simulation.

Conservatism:

Bonferroni method has the largest p values and thus most conservative methods, followed by the Holm (1979), Hochberg (1988), and Hommel (1988) methods. The Bonferroni and Holm (1979) methods shows the lowest Type I error, whereas the Hochberg (1988) and Hommel (1988) methods allowed more error but are still conservative when ρ (correlation) exceeded 0.5.

Holm procedure is a closed testing procedure in which each intersection hypothesis is tested using a global test based on the Bonferroni procedure. Holm procedure rejects the global hypothesis if and only if the Bonferroni procedure does and therefore the conclusions regarding the conservative nature of the Bonferroni procedure also apply to the Holm procedure (Dmitrienko et al, 2009).

Hochberg procedure uses the same criterion for each hypothesis as does the Holm procedure but tests hypotheses with larger p values first. Consequently this procedure will test and possibly reject hypotheses not examined by the Holm procedure while rejecting the same hypotheses that are rejected by the Holm procedure ((Dunnett & Tamhane, 1992; Hochberg, 1988; Olejnik et al, 1997). In most real-life cases, the conclusions from the two methods i.e. Holm & Hochberg will rarely differ.

Power:

Holm procedure is more powerful than Bonferroni method because the bound for this method sequentially increases whereas the Bonferroni bound remains fixed. Holm procedure is at least as powerful as Bonferroni because. Statistical power is gained by sequentially increasing the criterion for statistical significance. Because any hypothesis rejected by the original Bonferroni procedure will also be rejected by the Holm procedure, the latter procedure cannot have lower power for an individual hypothesis test. However, Holm claims that in actual practice the gain in power with his procedure as compared to Bonferroni is non negligible because orthant α/(c-i-1) is much larger than α/k for many values of i (Olejnik et al, 1997).

Any hypothesis rejected by Holm's procedure will always be rejected by Hochberg's procedure (Dunnett & Tamhane, 1992; Hochberg, 1988). However, the power differences tend to be negligible (Olejnik et al., 1997). Hochberg procedure is uniformly more powerful than the Holm procedure (Hochberg, 1988) but, on the other hand, it is uniformly less powerful than the Hommel procedure (Hommel, 1989). However, due to the independence assumption required by Hochberg, the Holm procedure may be the best choice if independence of tests is not certain. The criterion used by the Holland and Copenhaver procedure is slightly larger than Holm procedure thus leading to slightly greater power for an individual hypothesis test (Olejnik et al, 1997).

Hommel method is uniformly more powerful than Holm procedure because the Simes test is uniformly more powerful than the global test based on the Bonferroni procedure (Dmitrienko et al, 2009). For n>2, there are situation where Hommel reject and Hochberg does not reject (Hommel, 1989). Hommel procedure rejects more hypotheses than either the Rom or the Holland-Copenhaver procedure; however the difference in the number of tests rejected is very small. Hochberg and Hommel procedure are more powerful but they are known to have the desired FWE only for independent test (Hommel, 1989). Rom

gives slightly higher critical p-value that can be used with Hochberg's procedure, making it somewhat more powerful.

Holm's procedure is least powerful method, because it is based on the Bonferroni inequality. Rom procedure and Hommel procedure are more powerful than Hochberg's procedure due to the fact that sharp inequalities (or equalities) are used in both (i.e. Rom & Hommel) procedures; however, the power improvement is negligible compared to their complexities.

The increase in power for individual hypotheses tests provided by the Hommel and Rom procedures over the Hochberg approach is at best marginal with the Rom procedure having only a slight advantage over the Hommel (Dunnet and Tamhane, 1992; Olejnik, 1997).

Holland Copenhaver and Hochberg procedures provide power very close to that obtained by the Hommel and Rom procedures, particularly when the total number of hypotheses tested is not too large. If the numbers of false null hypotheses are large, Hochberg procedure might provide a better chance of detecting all of them than the Holland-Copenhaver procedure.

As the sample size increases, power of statistics increases but when the number of variables in a matrix increases, the probability of rejecting all of the non-null hypotheses decreases. All five of the enhancements are more sensitive than the original Bonferroni procedure in detecting all true nonzero relationships. The difference between the original Bonferroni procedure and the enhancements increased as the number of true nonzero relationships increased. Very small differences in statistical power are found among the five enhancements to the original Bonferroni procedure. The Holm procedure is having the lowest sensitivity in detecting all true nonzero relationships, whereas the Rom procedure has the greatest power. When all the correlations are nonzero, the Hochberg, Hommel, and Rom procedures had the same estimated power.

Because step up sequential multiple comparisons are based on the Simes equality, which assumes independence of comparisons, it is reasonable to suggest that dependence or correlation between the means of groups should affect the Type I error control and power (Zweifel, 2014).

In summary, the comparison of (Bonferroni, Holm, Holland, Hochberg, Hommel, Rom), Bonferroni procedure has the lowest percentage of rejections and Hommel procedure has the highest percentage of rejections whenever differences exist among the procedures. Overall, the SU procedures are little more powerful than the SD procedures. Within the SU procedures, whenever differences occurred, the Hommel procedure has slightly higher percentage of rejections than the Hochberg procedure. Within the SD procedure, whenever difference occur, Holland procedure having a slightly higher percentage of rejection than Holm procedure (Olejnik et al, 1997).

Confidence Interval:
All the methods are step wise methods except Bonferroni so confidence interval cannot be obtained by any of the method so comparison is not possible with respect to Confidence Interval.

Simulation Study:
This section discuss results regarding tests to be reported as adjusted p values such that, if the adjusted p value for an individual hypothesis is less than the chosen significance level α, then the hypothesis is rejected with FWE not more than α. It includes Bonferroni procedure and modification of that procedure by Holm, Holland & Copenhaver, Hommel, Hochberg and Rom.

As a concrete example, imagine that we have ten p values, and they are (in order from smallest to largest) as follows: 0.002, 0.0054, 0.007, 0.008, 0.009, 0.0094, 0.012, 0.015, 0.028, and 0.067.

We will compare probability with critical value based on Bonferroni method and modification of that procedure by Holm, Holland & Copenhaver, Hommel, Hochberg and Rom.

Table 1: Rejection criteria according to different available Tests

No.	Prob.	Bonferroni	Holm	Holland & Copenhaver	Hommel	Hochberg	Rom
1	0.002	0.005	0.005	0.005116197	0.025	0.005	0.005115
2	0.0054	0.005	0.005556	0.005683045	0.025	0.005556	0.005681
3	0.007	0.005	0.00625	0.006391151	0.025	0.00625	0.006388
4	0.008	0.005	0.007143	0.007300832	0.025	0.007143	0.0073
5	0.009	0.005	0.008333	0.008512445	0.025	0.008333	0.008505
6	0.0094	0.005	0.01	0.010206218	0.025	0.01	0.0102
7	0.012	0.005	0.0125	0.012741455	0.025	0.0125	0.0127
8	0.015	0.005	0.016667	0.016952428	0.025	0.016667	0.016875
9	0.028	0.005	0.025	0.025320566	0.025	0.025	0.025
10	0.067	0.005	0.05	0.05	0.025	0.05	0.05

Table 2: Hypotheses Rejection by all these multiple comparison procedure

No.	Bonferroni	Holm	Holland & Copenhaver	Hommel	Hochberg	Rom
1	Reject	Reject	Reject	Reject	Reject	Reject
2	Accept	Reject	Reject	Reject	Reject	Reject
3	Accept	Accept	Accept	Reject	Reject	Reject
4	Accept	Accept	Accept	Reject	Reject	Reject
5	Accept	Accept	Accept	Reject	Reject	Reject
6	Accept	Accept	Accept	Reject	Reject	Reject
7	Accept	Accept	Accept	Reject	Reject	Reject
8	Accept	Accept	Accept	Reject	Reject	Reject
9	Accept	Accept	Accept	Accept	Accept	Accept
10	Accept	Accept	Accept	Accept	Accept	Accept

From simulation Study also, we can see that Holm procedure is more powerful than Bonferroni method because the bound for this method sequentially increases whereas the Bonferroni bound remains fixed. Any hypothesis rejected by the original Bonferroni procedure will also be rejected by the Holm procedure; the latter procedure cannot have lower power for an individual hypothesis test. Any hypothesis rejected by Holm's procedure will always be rejected by Hochberg's procedure. Hochberg procedure is uniformly more powerful than the Holm procedure but, on the other hand, it is uniformly less powerful than the Hommel procedure. The criterion used by the Holland and Copenhaver procedure is slightly larger than Holm procedure thus leading to slightly greater power for an individual hypothesis test. Hommel procedure rejects more hypotheses than either the Rom or the Holland-Copenhaver procedure; however the difference in the number of tests rejected is very small. Rom gives slightly higher critical p-value that can be used with Hochberg's procedure, making it somewhat more powerful.

Table 3: Comparison of Multiple Comparison procedure

Test	SS/SW	Based on	Modification of	Remarks
Bonferroni	SS	Bonferroni inequality	——	Planned contrasts, both simple and complex.
Holm (1979)	SD	Bonferroni inequality	Bonferroni	comparisons are not independent
Holland (1987)	SD	Sidak inequality	Bonferroni	Positive orthant dependence
Hommel (1988)	SU	Simes inequality	Holm	When comparisons are independent
Hochberg (1988)	SU	Simes inequality	Holm	When comparisons are independent
Rom (1990)	SU	Simes Inequality	Hochberg	When comparisons are independent

REFERENCES:

1. Abdi, H. (2010). Holm's Sequential Bonferroni Procedure. In N.J. Salkind, D.M., Dougherty, & B. Frey (Eds.): Encyclopedia of Research Design, Thousand Oaks, CA: Sage. pp 573-577.
2. An, Q., Xu, D. & Brooks, P. (2013). Type I Error Rates and Power of Multiple hypothesis testing procedures in factorial ANOVA. Multiple Linear Regression Viewpoints. Vol. 39. No.2 .pp 1-16.

3. Dunnett, C.W. and Tamhane. A. C.(1992). A step-up multiple test procedure. Journal of the American Statistical Association 87, 162-170.

4. De Muth, J. (2006). Basic Statistics and Pharmaceutical Statistical Applications. Second Edition.CRC Press.2nd edition.

5. Dmitrienko, A.,Tamhane, A. & Bretz F. (2009), Multiple Testing Problems in Pharmaceutical Statistics. CRC press.

6. Einot, I. & K. R. Gabriel. (1975). A Study of the Powers of Several Methods of Multiple Comparisons. Journal of American statistical association, 70, pp 574-583.

7. Ekenstierna, M. (2004). Multiple comparison procedures based on marginal pvalues. Uppsala University Department of Mathematics Project Report No: 2004(12), 1-26.

8. Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75, 800–803.Howell, D. (2007), “Statistical Methods For Psychology,”7th edition.

9. Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics Scand. J. Vol 6, pp 65–70.

10. Hochberg, Y.& Benjamini, Y. (1990). More powerful procedures for multiple significance testing. Statistics in Medicine, Vol 9, pp 811–818.

11. Holland, B.S. & Copenhaver, M.D. (1987) “An improved sequentially rejective Bonferroni test procedure.” Biometrics, 43: 417-423.

12. Hommel G (1989) A comparison of two modified Bonferroni procedures. Biometrica 76,624-625.

13. Howell, D. C. (2007). Statistical Methods For Psychology. 7th edition. Pacific Grove, CA: Duxbury Press.

14. Li, H. (2009). Multiple comparisons under unequal variances and its application to dose response studies. A Dissertation Submitted to the Graduate College of Bowling Green State University.

15. McHugh, M. M. (2011). Multiple comparison analysis testing in ANOVA. Biochemia Medica, Vol. 21, No. 3, pp 203-209.

16. Newman, M.C. (1995). Quantitative Methods in Aquatic Ecotoxicology. Lewis Publishers, CRC press.

17. Olejnik, S., Li J, Supattathum S & Huberty S (1997). Multiple Testing and Statistical Power with Modified Bonferroni Procedures. Journal of Educational and Behavioral Statistics. Vol 22, pp 389-406.

18. Ott, R. & Longnecker, M. (2010). An Introduction to Statistical Methods and Data Analysis. Cengage Learning, 6th edition.

19. Schochet, P. Z. (2008). Guidelines for multiple testing in impact evaluations of educational interventions. Mathematica Policy Research, Inc., 1-34

20. Rom, D.M., 1990. A sequentially rejective test procedure based on a modified Bonferroni inequality. Biometrika 77, 663–665.

21. Sidak, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. JASA, 62, pp 626-633.

22. Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. Biometrika, Vol 73, pp 751–754.

23. Tamhane, A. C., Liu, A. W. and Dunnett, C. W. (1998). A generalized step up-down multiple test procedure. The Canadian Journal of Statistics, 26,353-363.

24. Tamhane, A. and Dunnett, C. (1999). Stepwise multiple test procedures with biometric applications. Statistical planning and inference, 82, pp 55–68.

25. Toothaker L. E. (1993). Multiple Comparisons procedures. Newbury Park, CA: Sage.

26. Wright, S. P. (1992). Adjusted p-values for simultaneous inference. Biometrics. Vol. 48, pp1005–1013.

27. Zweifel, M.(2014). The Power and Type I Error Rate of Holm's Procedure When the Assumptions of Normality and Variance Homogeneity are Violated. Open Access Theses and Dissertations from the College of Education and Human Sciences.